



O USO DE REDES NEURAI E REGRESSÃO LINEAR MÚLTIPLA NA ENGENHARIA DE AVALIAÇÕES: DETERMINAÇÃO DOS VALORES VENAI DE IMÓVEIS URBANOS

Marisa Baptistella

Universidade Federal do Paraná – UFPR
Programa de Pós-Graduação em Métodos Numéricos em Engenharia
CP: 19081; Curitiba, PR; CEP: 81531-990
e-mail: marisabaptis@yahoo.com.br

Maria Teresinha Arns Steiner

Universidade Federal do Paraná – UFPR
Departamento de Matemática - Programa de Pós-Graduação em Métodos Numéricos em
Engenharia - CP: 19081 – Curitiba, PR; CEP: 81531-990
e-mail: tere@mat.ufpr.br

Anselmo Chaves Neto

Universidade Federal do Paraná – UFPR
Departamento de Matemática - Programa de Pós-Graduação em Métodos Numéricos em
Engenharia - CP: 19081 – Curitiba, PR; CEP: 81531-990
e-mail: anselmo@est.ufpr.br

RESUMO

O presente trabalho tem por objetivo a utilização da técnica de Redes Neurais Artificiais na estimação dos valores venais de imóveis urbanos da cidade de Guarapuava, PR. Para tanto, utilizou-se dados do Cadastro Imobiliário fornecido pelo setor de Planejamento da Prefeitura Municipal. O modelo inicial foi composto por treze variáveis/atributos do cadastro: bairro, setor, pavimentação, esgoto, iluminação pública, área do terreno, pedologia, topografia, situação, área edificada, tipo, estrutura e conservação. A técnica da Análise das Componentes Principais foi usada para reduzir e transformar as variáveis originais em nove fatores. As Redes Neurais Artificiais desenvolvidas foram do tipo *feedforward*, utilizando o algoritmo de treinamento *Levenberg-Marquardt*, com uma camada oculta. Os resultados da amostra de dados completa foram comparados com os resultados obtidos dividindo-se a mesma em pequenos grupos, sendo que os obtidos nas pequenas amostras foram superiores aos primeiros.

PALAVRAS CHAVE: Redes Neurais Artificiais, Análise Multivariada e Engenharia de Avaliações. Área: Mataheurística (MH).

ABSTRACT

The goal of the present work is to use the Artificial Neural Networks technique in the market values estimation for urban real properties in the city of Guarapuava, PR. For this purpose, data from official register real estate were used, provided by the Planning Sector at the City Hall. The initial model encompassed thirteen variables/attributes from the register: quarter, sector, paving, drain, public illumination, land area, pedology, topography, status, built area, type, structure and conservation state. The Principal Component Analysis technique was used in order to reduce and to transform the original variables into nine factors. The Artificial Neural Networks developed were of feedforward type, using the Levenberg Marquardt training algorithm with one hidden layer. The results obtained of the complete data sample were compared to the results obtained by dividing the sample into small groups and they were better for the smaller samples, which performed better than the previous one.

KEYWORDS: Artificial Neural Networks, Multivariate Analysis and Engineering of Evaluations. Main area: Metaheuristics.



1. Introdução

A Engenharia de Avaliações, cada vez mais, vem evoluindo no Brasil, através da ação de inúmeros profissionais e pesquisadores que se dedicam ao estudo e pesquisa de problemas da área, bem como a divulgação das técnicas de avaliação de imóveis.

Segundo DANTAS (2003), a Engenharia de Avaliações é uma especialidade da engenharia que reúne um conjunto amplo de conhecimentos da área de engenharia e arquitetura, bem como de outras áreas das ciências sociais, exatas e da natureza, com o objetivo de determinar tecnicamente o valor de um bem, de seus direitos, frutos e custos de produção.

MENDONÇA et al. (1998) citam as situações mais frequentes em que a avaliação técnica se faz necessária, dividindo-as em três âmbitos. No âmbito particular, os autores citam o uso da avaliação de imóveis na compra e venda de imóveis em que os interessados desejem ter idéia precisa do bem em questão; na reavaliação de ativos de empresas para efeito de aumento de capital; pelas empresas para atendimento ao contido na Lei das Sociedades Anônimas; em divisões, heranças e meações, e nas discussões acerca de lançamentos de impostos.

No âmbito público, é citado o uso da avaliação de imóveis para fins de compra e privatização e no cálculo de valores para lançamentos de impostos. Finalmente, no âmbito judicial é utilizada nas discussões entre pessoas físicas ou jurídicas que envolvam valores de imóveis, frequentes em ações demarcatórias, possessórias e indenizatórias e, também, nas discussões acerca de indenizações por desapropriações ou servidões de passagem.

Os métodos estatísticos são os procedimentos de modelagem matemática mais utilizados para a avaliação imobiliária. O modelo de Regressão Linear Múltipla é o preferido dos avaliadores, por ter se mostrado bastante eficiente, embora fatores tais como: a complexidade dos modelos, dificuldades de implementação, excesso de variáveis envolvidas e desconhecimento da relação entre estas variáveis, possam comprometer a análise dos resultados. A escolha de Redes Neurais Artificiais, para a estimação do valor de um imóvel, apresenta como vantagens a não necessidade de linearidade entre as variáveis, se for o caso, e a facilidade de implementação.

2. Descrição do Problema e Obtenção dos Dados

Os dados referentes a esta pesquisa foram obtidos junto ao setor de planejamento da Prefeitura Municipal de Guarapuava. Para a obtenção dos dados foi aplicado-se um filtro ao banco de dados, separando para a análise 300 imóveis residenciais (casas e apartamentos), com valores venais entre R\$ 40.000,00 a R\$ 200.000,00 localizados na área urbana.

As variáveis escolhidas foram, basicamente, as constantes no Boletim de Informações Cadastrais, sendo elas: localização (bairro), setor, pavimentação, iluminação pública, esgoto, área do terreno, pedologia, topografia, situação, área edificada, tipo, estrutura, conservação e valor venal. As variáveis água e rede elétrica foram desconsideradas do modelo, pois todos os imóveis da amostra estão contemplados com estes itens.

A amostra foi reduzida de 300 para 256 unidades, pois eliminou-se os imóveis repetidos como, por exemplo, um prédio com vários apartamentos com as mesmas características, foi considerado apenas um apartamento.

3. Redes Neurais

As Redes Neurais Artificiais (RNA's), amplamente utilizadas pela comunidade científica, têm o seu funcionamento inspirado na estrutura biológica do cérebro humano e podem ser entendidas como conjuntos bem estruturados de unidades de processamento, interligadas por canais de comunicação, cada qual tendo um determinado peso correspondente a um valor numérico. No decorrer do processo de "aprendizagem", os referidos pesos vão se ajustando de forma a atingir o objetivo pretendido que, no presente trabalho, é a previsão dos valores venais dos imóveis.

3.1 Algoritmo de *Levenberg-Marquardt*

Em uma RNA de Múltiplas Camadas, o algoritmo *back-propagation* padrão utiliza o método do gradiente decrescente no decorrer do processo de aprendizagem de forma a minimizar a função erro. Já o algoritmo de *Levenberg-Marquardt (LM)* utiliza uma aproximação pelo método de Newton; esta aproximação é obtida a partir do método de Gauss-Newton modificado. A atualização dos pesos através do método de Gauss-Newton é feita da seguinte forma (HAYKIN, 2001):

$$w(n+1) = w(n) - (J^T(n)J(n) + \delta I)^{-1} J^T(n) e(n)$$

onde n é o número da iteração; $w(n)$ é o peso na iteração n ; $J(n)$ é a matriz jacobiana do vetor do erro em relação ao vetor peso na iteração n ; δ é uma constante positiva pequena escolhida para assegurar que $(J^T(n)J(n) + \delta I)$ seja definida positivamente para todo n e $e(n)$ é o erro na iteração n .

4. Regressão Linear Múltipla

A técnica estatística da Regressão Linear Múltipla é usada para estudar a relação entre uma variável dependente e várias variáveis independentes explicativas; ela também é bastante aplicada pela comunidade científica. Em Engenharia de Avaliações geralmente trabalha-se com Modelos de Regressão Múltipla, tendo-se em vista a multiplicidade de fatores que interferem nos preços de um bem (DANTAS, 2003).

O Modelo Genérico Linear Geral é dado pela expressão a seguir (LEVINE et al., 2005) quando aplicado a uma amostra de tamanho n ,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

onde:

y_i = variável dependente ou explicada $i = 1, 2, \dots, n$, também chamada de resposta;

β_0 = intercepto ou termo independente de variável;

...

β_k = inclinação de Y em relação à variável X_k , mantendo constantes as variáveis X_1, X_2, \dots, X_{k-1}

ε_i = erro aleatório em Y , para a observação i , $i = 1, 2, \dots, n$.

A aplicação deste modelo requer que $\varepsilon_i \sim N(0, \sigma^2)$, ou seja, que os erros das observações i sejam Gaussianas, independentes, com média zero e variância constante.

Além disso, é necessário:

- Aplicar a Análise da Variância da Regressão que é uma técnica estatística usada para verificar se o ajuste de regressão existe;
- Avaliar o poder de Explicação do Modelo que verifica se o ajuste do modelo é adequado (está intimamente ligado à estatística F). O coeficiente de determinação múltipla que informa o poder de explicação do modelo em função das p variáveis independentes consideradas, dado por R^2 . Quando o ajuste é bom, o modelo explica boa parte da variação total e, consequentemente, o valor de R^2 é próximo de 1, ou seja, o coeficiente de determinação é uma medida da qualidade do ajuste.

5. Análise das Componentes Principais

A Análise das Componentes Principais é uma das técnicas da Análise Estatística Multivariada utilizada para investigar a estrutura de covariância em um conjunto de p variáveis correlacionadas. Para tanto, transforma-se o conjunto das p variáveis originais em um novo

conjunto com cardinal $k < p$ de variáveis não-correlacionadas chamadas de Componentes Principais, tendo propriedades especiais em termos de variâncias.

Assim, a Análise das Componentes Principais explica a estrutura de variância e covariância de um vetor aleatório através de poucas combinações lineares das variáveis originais. O objetivo geral consiste, tanto em reduzir os dados como em facilitar a interpretação, pois trata-se de uma transformação ortogonal de eixos, tornando as novas variáveis (combinações lineares) não correlacionadas (JOHNSON e WICHERN, 2002) e nas direções com maior variabilidade.

A reprodução da variabilidade total do sistema requer as p variáveis, porém, freqüentemente a maior parte dessa variabilidade pode ser explicada por um número pequeno $k < p$, de componentes principais. Neste caso, existe praticamente a mesma quantidade de informações nas k componentes principais do que nas p variáveis originais.

5.1 Componentes Principais da Amostra

Geralmente a estrutura de covariância, resumida na matriz de covariância Σ é desconhecida, então a obtenção das componentes principais é feita a partir de estimativas, da matriz de covariância, S , ou da matriz de correlação, R . Então, obtém-se as estimativas dos elementos da estrutura de covariância do vetor aleatório \underline{X} , ou seja, os autovalores e os correspondentes autovetores e constrói-se as componentes principais amostrais. As propriedades das componentes principais se mantêm e são obtidas com base em estimadores.

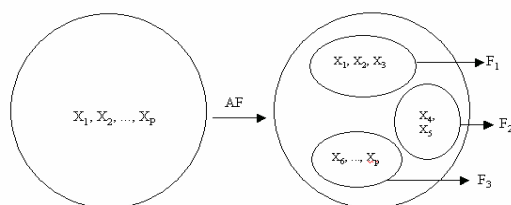
A obtenção das componentes principais com base nas informações da matriz de correlação é preferida, pelo fato de se conseguir eliminar o efeito de escala nos valores das componentes do vetor de variáveis originais X . Como é bem conhecida, a matriz de correlação é uma matriz de covariância, mas de variáveis padronizadas. Assim, consegue-se eliminar a influência da escala na magnitude das variáveis (JOHNSON e WICHERN, 2002).

Os autovetores definem as direções da máxima variabilidade e os autovalores especificam as variâncias. Quando alguns autovalores são muito maiores que os demais, significa que a maior parte da variância total pode ser explicada por um número muito menor de componentes principais, do que as p dimensões do vetor \underline{X} .

5.2 Análise Fatorial

A Análise Fatorial é uma técnica de Análise Multivariada que objetiva explicar as correlações existentes entre um conjunto grande de variáveis em termos de um conjunto de poucas variáveis aleatórias não observáveis, denominadas fatores. A Análise Fatorial pode ser melhor visualizada no esquema da figura 5.1 que se segue:

Figura 5.1. Esquema de Agrupamento de Variáveis em Fatores (AF)



Fonte: JOHNSON e WICHERN (2002)

Quanto mais fortes forem as correlações entre algumas variáveis dentro o grupo inicial, mais nítida é a visualização do fator gerado. Variáveis agrupadas num mesmo fator possuem, portanto, alta correlação, enquanto que variáveis de fatores distintos possuem baixa correlação.

5.3 Rotação dos Fatores

Essa técnica é empregada para facilitar a interpretação dos fatores. A rotação consiste em girar os eixos em um ângulo q , oferecendo uma nova estrutura para os pesos de tal forma que cada variável tenha peso alto em um único fator e pesos mais baixos ou médios nos demais fatores. A rotação não produz uma estrutura visível quando o número de fatores é maior que 2, ou seja, $m > 2$. A técnica de rotação de fatores mais utilizada foi proposta por Kaiser e recebeu o nome de Rotação Varimax ou Normal Varimax.

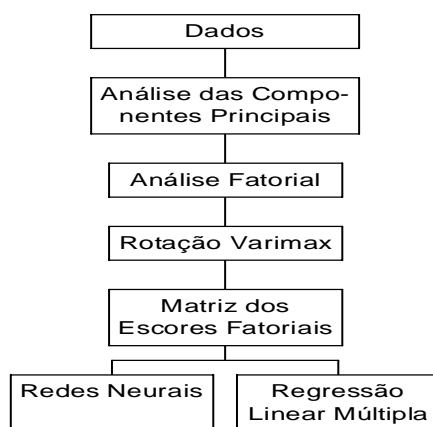
5.4 Escores Fatoriais

Os fatores são variáveis latentes, ou seja, não observáveis. Seus valores, porém, podem ser estimados por meio de regressão. Os valores estimados dos fatores são denominados escores. Se ocorrer rotação, os escores são obtidos a partir das cargas fatoriais que sofreram rotação. Não obstante, as fórmulas não distinguirão entre as situações em que ocorreu rotação daquelas em que não ocorreu, uma vez que estas fórmulas não são alteradas pelas rotações (JOHNSON e WICHERN, 2002).

6. Implementação das Técnicas para a Resolução do Problema

Após o levantamento dos dados, implementou-se as técnicas propostas neste trabalho, conforme fluxograma 6.1 apresentado a seguir.

Fluxograma 6.1. Implementação das Técnicas



A avaliação estatística das características dos dados foi realizada por meio de uma Análise das Componentes Principais. Esta operação teve como propósito avaliar a importância relativa das variáveis que compõem a amostra de dados. Este procedimento pode ser adotado em situações onde o número de variáveis (entradas) do modelo é grande e as mesmas são altamente correlacionadas. Esta análise possibilita a redução da dimensão dos vetores de entrada. Operacionalmente trabalhou-se, na Análise das Componentes Principais, com o *software Estatística 5.0*.

Aplicando-se a técnica das Componentes Principais ao conjunto de dados originais composto das treze variáveis já especificadas percebeu-se que um modelo com apenas nove componentes principais seria adequado para representar a estrutura de covariância inicial. E, ainda, obteve-se com este modelo apenas 18,69% de perda de informações. Nesse modelo considerou-se autovetores cujos autovalores são maiores que 0,75. Na tabela 6.1 observa-se a proporção da variância explicada pelos autovalores ordenados de forma decrescente para cada uma das treze variáveis.

Tabela 6.1. Autovalores em Ordem Decrescente de Magnitude

	Autoval	% total Variância	Cumul. Autoval.	Cumul. %
1	1,858557	14,29659	1,85856	14,29659
2	1,511854	11,62965	3,37041	25,92624
3	1,296738	9,97491	4,66715	35,90115
4	1,208301	9,29463	5,87545	45,19577
5	1,100844	8,46803	6,97629	53,66380
6	1,044903	8,03772	8,02120	61,70152
7	0,896430	6,89561	8,91763	68,59713
8	0,831275	6,39442	9,74890	74,99155
9	0,821521	6,31939	10,57042	81,31094
10	0,727890	5,59916	11,29831	86,91010
11	0,653833	5,02948	11,95215	91,93960
12	0,555787	4,27529	12,50793	96,21490
13	0,492067	3,78513	13,00000	100,00000

Assim, a aplicou-se a Análise Fatorial com carregamentos obtidos pelo método das Componentes Principais. Refinou-se o processo com a Rotação Varimax e os pesos dos fatores selecionados estão na tabela 6.2. Desta forma, analisando os fatores com pesos contrastados pode-se interpretar os fatores.

O fator 1 tem carregamento alto na variável bairro e esgoto, logo o identificamos como “localização”; o fator 2 depende mais fortemente da variável conservação; o fator 3 tem carregamento alto nas variáveis iluminação e pavimentação, que será identificado como “serviços urbanos”; o fator 4 tem carregamento alto nas variáveis tipo e área do terreno, identificado como “característica do imóvel”; o fator 5 depende fortemente da variável setor; o fator 6 depende da variável área edificada; o fator 7 depende da variável pedologia; o fator 8 depende da variável topografia e o fator 9 tem carregamento forte na variável situação.

Tabela 6.2. Matriz dos Pesos com Rotação Varimax

	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9
BAIR	-0,769878	0,118636	0,088964	-0,024202	0,097856	-0,268058	0,011785	0,067484	0,065238
SETO	0,050972	-0,002536	0,040935	0,008548	-0,935737	-0,080292	-0,046661	0,039636	0,015058
PAVI	-0,126354	0,093351	-0,714315	0,066778	0,321259	-0,050776	-0,090172	0,049206	0,016491
ESGO	0,794818	0,157260	0,080590	0,190432	0,016297	-0,117385	-0,043827	0,081426	0,135939
ILUM	-0,108389	0,050869	0,830505	0,043004	0,170622	0,015022	-0,083221	-0,015309	0,053253
ATER	-0,274650	0,017026	-0,001866	-0,750615	-0,0100,,	0,316163	-0,086787	0,094259	0,083614
PEDO	-0,044983	0,057990	-0,008796	0,067410	0,041361	0,008730	0,983214	0,006888	0,032191
TOPO	-0,010974	0,024510	0,052197	0,094515	0,037077	0,003995	-0,007034	-0,986434	-0,011558
SITU	0,051411	-0,035202	0,034461	0,031832	-0,014326	0,062095	0,032251	0,011690	0,977704
AEDI	0,112770	0,110913	0,062490	0,047113	0,087366	0,876510	0,009610	-0,007188	0,063956
TIPO	0,041138	-0,032906	-0,004067	0,809722	0,006085	0,252677	0,002514	-0,047450	0,118306
ESTR	0,235871	-0,579661	0,072696	-0,431292	0,130118	-0,295016	-0,113405	-0,044759	0,136962
CONS	0,107314	0,912569	0,001780	-0,124559	0,056250	0,028308	0,021534	-0,040724	0,009897

Finalmente, aplicando o método dos mínimos quadrados encontra-se a matriz dos escores fatoriais de ordem $n \times 9$. Esta matriz passa a representar os dados.

6.1 Modelo utilizando Redes Neurais

Para as RNA's foi utilizado o *software Matlab 6.5*. A topologia empregada foi a *MLP (MultiLayer Perceptron) feedforward* com uma camada de entrada, consistindo de nove nós,

conforme as variáveis obtidas na Análise das Componentes Principais; uma camada escondida, com um número de neurônios variando de zero a doze; e uma camada de saída, com um único neurônio que fornecerá o valor venal do imóvel. Para as camadas oculta e de saída foi utilizada a função de ativação não-linear sigmoidal (logsig), que assume um intervalo contínuo de valores entre zero e um e é definida como uma função estritamente crescente, que exibe um balanceamento adequado entre comportamento linear e não-linear.

O treinamento da rede *MLP* foi feito utilizando-se o algoritmo de *LM*. De acordo com HAYKIN (2001), uma rede *MLP* treinada com o algoritmo de retropropagação pode ser vista como uma forma prática para realizar um mapeamento não-linear de entrada-saída de natureza geral.

O desempenho da rede neural foi avaliado por meio do erro quadrático médio (*MSE* – *Mean Squared Error*), que é a média do somatório dos quadrados dos erros de cada caso, tanto do conjunto de treinamento, quanto do conjunto de testes. É normal se adotar a raiz quadrada do erro quadrático médio (*rmse* – *root mean squared error*) no cálculo do erro da rede.

Na continuidade, a amostra de dados foi dividida em dois grupos. O primeiro grupo chamado de Conjunto de Treinamento (P), com 66% da amostra (170 imóveis), foi utilizado para treinar a rede e o segundo, chamado de Conjunto de Teste (S), com 34% (86 imóveis), foi utilizado para testar a rede neural treinada. Nesta divisão tomou-se o cuidado de manter a mesma proporção de valores dos imóveis nos dois conjuntos.

Foi escolhida uma nomenclatura para as redes de forma a representar, pela ordem, as seguintes características: número de parâmetros de entrada, número de neurônios na camada escondida, número de neurônios na camada de saída (sempre igual a um, neste trabalho) e número de iterações para o treinamento. Desta forma, a denominação das redes deste trabalho começa com um número correspondente à quantidade de variáveis de entrada, seguido da letra E; seguida do número de neurônios na camada oculta, seguido da letra N; seguido do número de neurônios na camada de saída, seguido também pela letra N e, finalmente, o número de iterações no treinamento. Por exemplo, a rede 9E8N1N50 é uma rede com 9 Entradas, 8 Neurônios na camada oculta, 1 Neurônio na camada de saída que foi treinada com 50 iterações.

Cada conjunto foi treinado aproximadamente 50 vezes, variando-se o conjunto de pesos iniciais, e o critério de parada utilizado foi o número de iterações (épocas) para cada treinamento. As simulações que apresentaram melhores resultados, são mostradas no quadro 6.1 a seguir.

É possível observar neste quadro que as redes apresentam para os conjuntos de treinamento, valores do *rmse* no intervalo de 2,81% a 14,89%. Estes valores para os conjuntos de testes, ficaram no intervalo entre 10,40% e 16,41%. A rede com 5 neurônios na camada oculta apresentou o melhor resultado de teste (*rmse* = 10,40%, ou seja, 89,60% de acerto), ou seja, uma melhor capacidade de generalização.

6.2 Modelo Utilizando Análise de Regressão Múltipla

O *software* utilizado para a construção da tabela de dados e aplicação do Modelo de Regressão Múltipla foi o *Excel*, com validação dos resultados através do *software Statistica v.5*. Ajustou-se o Modelo de Regressão Linear Múltipla para os dados observados, a partir da matriz obtida após a Análise das Componentes Principais. No ajuste do modelo verificou-se que as variáveis: setor, pedologia e topografia, não são significativamente importantes devido aos seus valores-*p* serem maiores que 0,05; assim, estas variáveis foram excluídas do modelo. A tabela 6.3 apresenta as variáveis utilizadas e os respectivos valores dos coeficientes.

Quadro 6.1. Resultado das Melhores Simulações

Rede	Erro Médio Quadrático <i>rmse (%)</i>	
	Treinamento	Teste
9E1N50	14,89%	16,41%
9E1N1N50	12,32%	14,82%
9E2N1N50	11,39%	15,39%
9E3N1N50	7,61%	13,15%
9E4N1N50	6,96%	12,50%
9E5N1N50	6,28%	10,40%
9E6N1N50	5,62%	12,09%
9E7N1N50	3,91%	13,94%
9E8N1N50	4,96%	11,13%
9E9N1N50	3,91%	13,81%
9E10N1N50	4,64%	12,88%
9E11N1N50	3,14%	12,23%
9E12N1N50	2,81%	15,81%

Tabela 6.3. Ajuste do Modelo Final de Regressão

Parâmetro	Coefficientes	Erro padrão	Estatística <i>t</i>	Valor- <i>p</i>
Interseção	90070,2	1800,4016	50,027852	6,586E-132
Localização	7724,55	1803,9283	4,2820725	2,64519E-05
Estrutura	15102,6	1803,9283	8,3720638	4,13014E-15
Serviços Urbanos	4649,08	1803,9283	2,5771964	0,010536985
Característica	-9013	1803,9283	-4,996315	1,10135E-06
Área Edificada	42448	1803,9283	23,530889	3,11186E-65
Situação	5028,31	1803,9283	2,787424	0,005722526

O coeficiente de determinação múltipla (R^2), que representa a proporção da variação em Y e é explicada através do conjunto de variáveis explanatórias selecionadas, apresentou um valor igual 0,7324, ou seja 73,24% da variação no valor venal pode ser explicado através da variação nas variáveis e, 26,76% dos valores dos imóveis, são explicados por outras variáveis que não constam no modelo, conforme pode ser observado na tabela 6.4.

Tabela 6.4. Resultado da Estatística

Coefficiente de Correlação	0,85581
R^2 (coef. Correlação ao quadrado)	0,73240
R^2 ajustado	0,72596

A equação de Regressão Linear Múltipla que descreve a relação entre o valor venal do imóvel e as seis variáveis independentes (localização, estrutura, serviços urbanos, característica, área edificada e situação) é:

$$\text{Valor Venal} = 90070,2 + 7724,55X_1 + 15102,6X_2 + 4649,08X_3 - 9013X_4 + 42448X_5 + 5028,31X_6$$

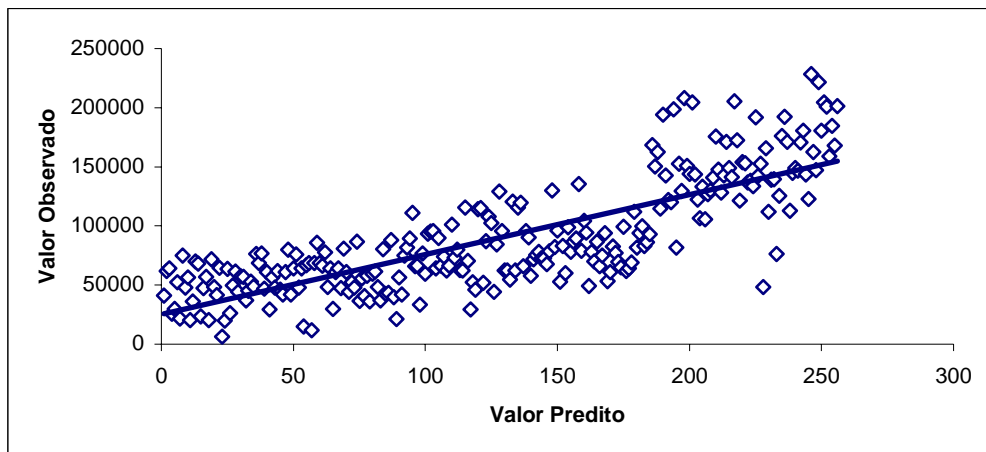
A Análise de Variância mostrou que a hipótese (H_0) de não haver regressão é rejeitada, isto é, o modelo é significativo a um nível de significância de 0,05. O valor crítico na distribuição F , com 6 e 249 graus de liberdade é, aproximadamente, 2,10. Uma vez que a estatística do teste é $F = 113,58 > F_s \cong 2,10$, ou, ainda, o valor- $p = 0,0000 < 0,05$, rejeita-se H_0 e conclui-se que pelo menos uma das variáveis explanatórias está relacionada com valor venal, conforme tabela 6.5.

Tabela 6.5. Análise da Variância do Ajuste do Modelo de Regressão

Fonte de Variação	Graus de Liberdade	Soma dos Quadrados	Quadrado Médio	Estatística F	Valor- p
Regressão	6	5,655E+11	9,425E+10	113,5841992	2,3E-68
Resíduo	249	2,066E+11	829810139		
Total	255	7,721E+11			

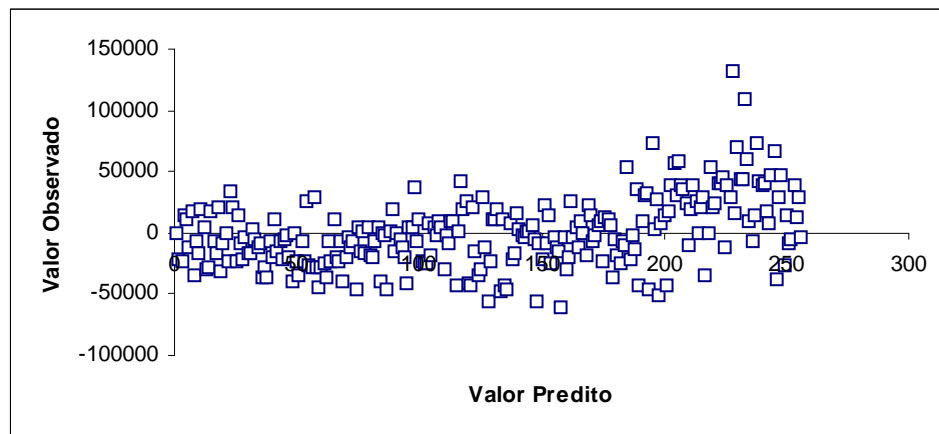
A figura 6.1 mostra que os valores preditos estão disposto em linha diagonal, indicando uma linearidade satisfatória, ou seja, as previsões se aproximam dos valores reais.

Figura 6.1. Valores Preditos x Valores Observados



A figura 6.2 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem, disposição esta, que indica que a suposição de variância constante está correta, ou a condição de homocedasticidade foi cumprida.

Figura 6.2. Resíduos x Valores Preditos



A estatística de Durbin-Watson (DW) mede a correlação entre cada resíduo e o resíduo correspondente ao período imediatamente antecedente àquele de interesse e é definido pela equação a seguir (LEVINE, 2005).

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

onde, e_i é o resíduo para o período de tempo i .

Quando os resíduos sucessivos são positivamente autocorrelacionados, o valor de D irá se aproximar de zero. Se os resíduos não forem correlacionados, o valor D estará próximo de 2. O teste de Durbin-Watson (DW), apresentou valor $D = 1,4993$, indicando não haver autocorrelação entre os resíduos.

Testou-se, ainda, a Gaussianidade dos resíduos através do teste de Kolmogorov-Smirnov que forneceu valor de $p = 0,06194$ ($> 0,05$), indicando que a distribuição dos resíduos é Gaussiana. O quadro 6.2 apresenta alguns valores preditos pela equação ajustada, os valores observados e a porcentagem de erro da predição. Assim, tem-se, por exemplo, que em 63 dos 256 casos (amostra de imóveis), ou seja, em 25% dos casos, a diferença entre o valor predito e o valor real (dados do Cadastro Imobiliário Municipal) difere em no máximo 10% e, da mesma forma, tem-se a interpretação para os demais casos deste quadro.

Quadro 6.2. Ocorrência de casos por faixa de erro no Modelo de Regressão

Erro (%)	Nº Casos	% de Casos
[0, 10]	63	25
(10, 20]	58	23
(20, 30]	41	16
(30, 40]	29	11
(40, 50]	17	7
(50, 60]	21	8
> 60	27	11
Total de Casos	256	100%

7. Conclusões

Este trabalho teve como objetivo comparar as técnicas de Redes Neurais Artificiais e Regressão Linear Múltipla para determinação de valor venal de imóveis na cidade de Guarapuava/PR. Utilizando os dados do Cadastro Imobiliário Municipal, utilizou-se a técnica de Redes Neurais Artificiais, aplicando o algoritmo *LM* com uma camada oculta à uma Rede de Múltiplas (três) Camadas e, também, a técnica estatística de Regressão Linear Múltipla. O modelo de Redes Neurais apresentou desempenho superior ao modelo estatístico de Regressão Linear Múltipla, pois apresentou um percentual de acerto de 89,60% contra 73,24% da regressão.

O algoritmo *LM* utilizado exige pouco tempo computacional, necessitando de um número relativamente pequeno de iterações (no trabalho utilizou-se 50 iterações), além do *software Matlab* que faz uso de Redes Neurais, apresentar uma forma acessível de interface com o usuário, o que facilita a aplicação. O uso da Regressão Linear Múltipla exige que o usuário tenha noções de álgebra, o que dificulta sua utilização e interpretação.

A Análise Multivariada aplicada reduziu o número de variáveis necessário ao cálculo, sem perda significativa de informação, ou seja, não prejudica a compreensibilidade do modelo obtendo variáveis não correlacionadas.



É importante observar, que outras variáveis poderiam melhorar o modelo de determinação de valores venais, uma vez que fatores como acessibilidade, distância do centro comercial, acesso a escolas, hospitais, supermercados também são importantes na determinação da aquisição de um imóvel e também para a sua valorização, desta forma estas poderiam ser levadas em consideração no modelo.

Referências:

- DANTAS, R. A. **Engenharia de Avaliações: uma introdução à metodologia científica**. 1.^a Ed. São Paulo: Pini, 2003.
- HAYKIN, S., **Redes Neurais: Princípios e Prática**. 2.^a edição, Porto Alegre: Bookman, 2001.
- JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 5.^a ed. Nova Jersey: Prentice Hall, Inc., 2002
- LEVINE, D.M; BERENSON, M.L.; STEPHAN, D. **Estatística: Teoria e Aplicações**. Rio de Janeiro: LTC, 2005.
- MENDONÇA, M.C.; SOLLERO FILHO, M.; CURI, E; AGUIAR, J.B.; QUEIROGA, H.S.; MAIA, E.A.; AQUINO, R.; RESENDE, O.; CANÇADO, J.M.M. **Fundamentos de Avaliações Patrimoniais e Perícias de Engenharia**. São Paulo: Editora Pini, 1998.